REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-03-

0383

1. AGENCY USE ONLY	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	26-08-2003	Final Performance Report: 01-06-2002 - 31-05-2003

4. TITLE AND SUBTITLE TeraWorld: Efficiently Accessing and Integrating Worldwide Geographic Data	5. FUNDING NUMBERS F49620-02-1-0270
6. AUTHORS Craig A. Knoblock and Cyrus Shahabi	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Southern California / Information Sciences Institute 4676 Admiralty Way, Suite 1001 Marina del Rey, CA 90292	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 4015 Wilson Blvd Arlington, VA 22203-1954	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES	
12A. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	

20031006 051

13. ABSTRACT *(Maximum 200 words)*

Geographic information such as satellite imagery, maps, and vector data play a critical role in military planning and intelligence. This type of information is being collected at an ever-increasing rate as new technology is developed, new commercial satellites are launched, and more and more information is being made available on the Internet. Unfortunately, the ability to efficiently access and integrate this type of data has not kept up with the availability of the information. In this project we purchased the hardware to build a multi-terabyte, worldwide geographic information system. The resulting system allows us to conduct research on a large-scale geographic information system, provide a significant advance in the military capabilities with respect to access to geographic data, and further the educational mission of the University by providing both research and teaching opportunities in the area of geographic information systems and information integration.

14. SUBJECT TERMS Geospatial information integration	15. NUMBER OF PAGES 8 pages
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
289-102

2003 Final Performance Report

TeraWorld: Efficiently Accessing and Integrating Worldwide Geographic Data

USAF, Air Force Office of Scientific Research
Award Number: F49620-02-1-0270
Period of Performance: 01-06-2002 – 31-05-2003

Craig A. Knoblock (PI)
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
Phone: 310-448-8786
Fax: 310-822-0751
Knoblock @isi.edu

Objectives of the Research Effort:

Geographic information such as satellite imagery, maps, and vector data play a critical role in military planning and intelligence. This type of information is being collected at an ever-increasing rate as new technology is developed, new commercial satellites are launched, and more and more information is being made available on the Internet. Unfortunately, the ability to efficiently access and integrate this type of data has not kept up with the availability of the information. We have been working on a joint project with DARPA and Special Operations Command to develop a system for rapidly accessing and integrating geographic data for any place in the world. The project makes extensive use of unclassified imagery, maps, and vector data from the National Imagery and Mapping Agency and exploits the latest research and commercial products for efficiently navigating and integrating geographic data. The system is currently being evaluated by SOCOM and SOCPAC for possible use for both military planning and intelligence. However, prior to this equipment grant, the system had only limited geographic coverage. In this project, we requested funds for the hardware to build a multi-terabyte, worldwide geographic information system. The resulting system allows us to conduct research on a large-scale geographic information system, provides a significant advance in the military capabilities with respect to access to geographic data, and furthers the educational mission of the University by providing both research and teaching opportunities in the area of geographic information systems and information integration.

Status of the Effort:

The equipment grant enabled us to build a multi-terabyte, worldwide geographic information system and populate the system with data covering the entire world. In particular, we purchased a Dell Storage Attached Network with eight high end servers and over four terabytes of storage. In addition, as proposed, we also purchased ten (10) laptop computers for developing and testing the remote applications that access the data on the servers and eight portable disk drives to provide access to portions of the data when network connectivity is not available. We then populated the resulting system with roughly two terabytes of NIMA imagery (CIB), maps (CADRG), vector (VPF, DNC and UTF), elevation (DTED), and points (GEONET Names). The products have different coverage for different parts of the world, but we now can provide some level of coverage for any part of the world and very detailed coverage for many parts of the world. In addition, we extended our software (HeraclesMaps) for provided a geospatial profile for a specified region to fully exploit the data available in our worldwide geographic information system.

Accomplishments:

Equipment Purchased

As planned we used this grant to purchase three main items (1) servers and attached storage, (2) the laptop computers, and (3) the portable disk drives. The servers and attached storage provide the core of the system where all of the data is hosted and the CPU intensive data processing is performed. The laptops run the end-user applications that send data requests to the servers. And the portable disk drives provide an approach to handling the case where the users do not have network connectivity, but still need to run the application.

Servers and Storage Attached Network	$174,744.83
10 Sony VAIO Laptop Computers	$30,212.31
8 Kanguru Portable Disk Drives	$3,652.72
Total Funds Expended:	$208,609.86
Total Funds Budgeted:	$208,841.00
Funds Remaining:	$231.14

The details of the servers and storage attached network are shown in the table below. This configuration consists of eight high-end servers, each of which is tuned for handling a specific type of geospatial data. This includes two vector servers, two image servers (for both imagery and maps), 2 database servers, and 2 web servers. It also includes the storage attached network (SAN), which provides 40 X 73 GB drives and an additional 650GB Maxtor drive.

The laptops that were purchased are all Sony VAIO Model V505 laptops along with the cases, adaptors, and wireless cards. These are very light, but powerful laptops that are ideal for running applications in the field.

Finally, the Kanguru drives are portable hard drives that hold between 120 and 250 GB of data. This is a substantial amount of storage, which allows us to download an entire portion of the world and then directly connect the laptops to the Kanguru drives to provide the data to run the applications when the laptops are disconnected.

Servers and Storage Attached Network

Vector Servers **Qty 2** 2650 Dual 1.8GHz Xeon Processors 2GB DDR memory 512KB L2 Cache (on die) 2 Gigabit Copper NIC Redundant Power Raid Cards with 2 internal Channels Rail Kit for Dell Rack 2 X 18 GB Drives Raid 1 (for OS and Log Files) 1 yr Service 7X24 4 hour response	Oracle Server **Qty 1** 6650 Quad 1.4 GHz Xeon Processors 2 GB DDR memory 512KB L2 Cache (on die) 2 Gigabit Copper NIC Redundant Power Raid Cards with 1 internal 1external channels Rail Kit for Dell Rack 2 X 18 GB drives Raid 1 (for OS and Log files) 1 yr Service 7X24 4 hour response
Image Servers ***Qty 2*** 2650 Dual 2.2GHz Xeon Processors 1GB DDR memory 512KB L2 Cache (on die) 2 Gigabit Copper NIC Redundant Power Raid Cards with 2 internal Channels Rail Kit for Dell Rack 2 X 18 GB Drives Raid 1 (for OS and Log Files) 1 yr Service 7X24 4 hour response	SQL Server **Qty 1** 2650 Dual 2.4 GHz Xeon Processors 1GB DDR memory 512KB L2 Cache (on die) 2 Gigabit Copper NIC Redundant Power Raid Cards with 2 internal Channels Rail Kit for Dell Rack 2 X 18 GB Drives Raid 1 (for OS and Log Files) 1 yr Service 7X24 4 hour response
SAN and Backup Unit ***Qty 1*** FC4500 w/ 1GB Cache 40 X 73 GB Drives Connectivity for 7 non redundant systems 16 Port Switch Navisphere Software Access Logic Analyzer Implementation for up to 7 servers PV128T Tape Backup 1 LTO Drive 20 Slot Library Fibre interface ArcServe Software for SAN Media LTO (20 tapes) Maxtor MaxAttach 650GB Drive	WEB Servers **Qty 2** 2650 Dual 1.8 GHz Xeon Processors 1GB DDR memory 512KB L2 Cache (on die) 2 Gigabit Copper NIC Redundant Power Raid Cards with 2 internal Channels Rail Kit for Dell Rack 2 X 18 GB Drives Raid 1 (for OS and Log Files) 1 yr Service 7X24 4 hour response

Research Supported

We are using the hardware to support research on the following topics:

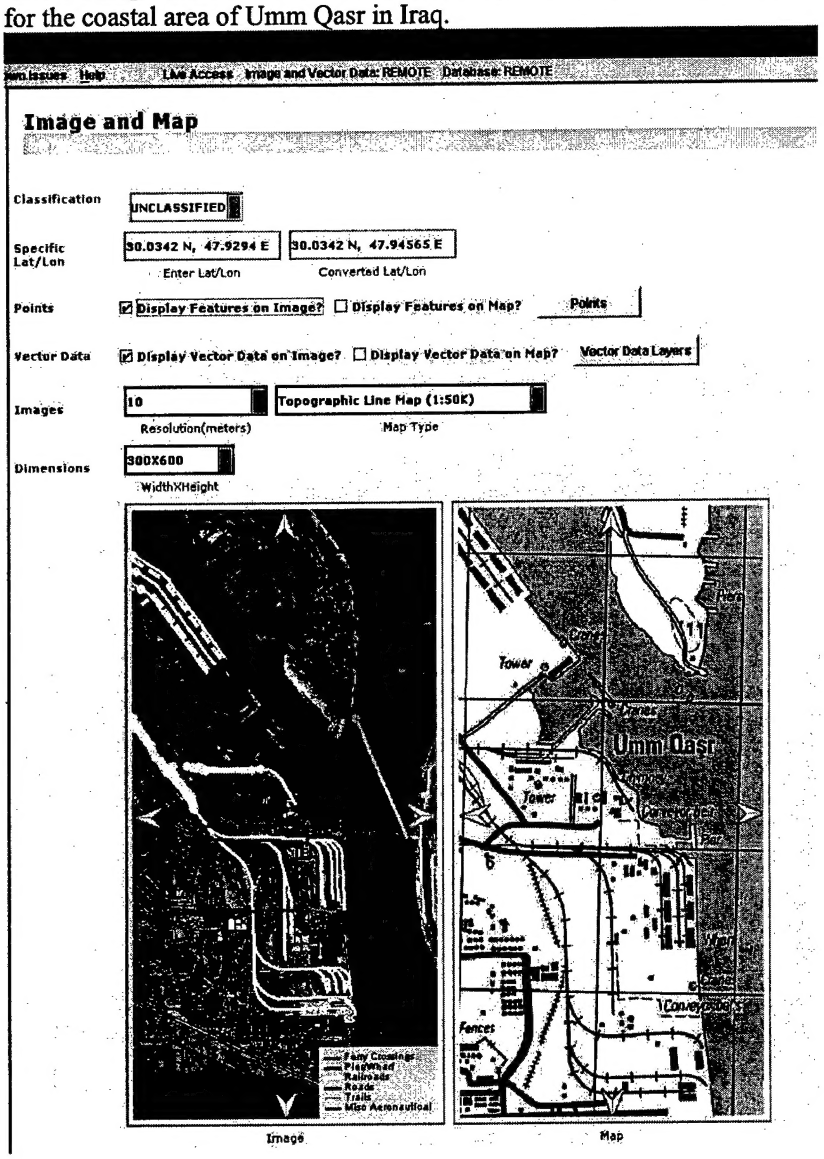
Formatted: Bullets and Numbering

- First, we are investigating the issues in building a large-scale, multi-terabyte geographic information system. We have access to satellite imagery, maps, and vector data for the entire world and the hardware allowed us to build, experiment, and deploy the system. We are developing algorithms and techniques to provide real-time support for accessing and integrating the geospatial data. For example, for efficient access to vector data we preprocessed the data and stored it into a commercially available spatial database. Similar issues arise in efficiently processing other types of geospatial data, such as digital elevation maps and imagery that is taken at different times. The digital elevation maps provide terrain elevations and are used to compute the field of view from an arbitrary location and imagery that is taken at different times can be used to determine the changes within an area over time. The data to perform these different types of analyses exists, but the challenge is how to efficiently organize, access, and process the data so that the information can be used effectively.
- Second, we are investigating architectures for distributed and disconnected access to the geographic information system. An important aspect of the problem for Special Operations Forces and the military in general is that they must continue to operate when they are not connected to the network or when they have a very low bandwidth connection. We are developing a distributed architecture that allows the user to specify a geographic region of interest and the system would then process the data for that region and store the required information on a portable disk. We are using the laptops and Kanguru drives to provide a complete system that will function seamlessly even when the client computers are disconnected from the network.
- Third, we are investigating the integration of open source data with the geographic data sources and the equipment allows us to demonstrate the scalability of this work. More specifically, we are using wrapper technology that we developed to extract semistructured data from open sources and integrate the extracted data with maps, imagery, and vector data. For example, we can extract worldwide address data from online white pages and yellow pages, geocode the addresses, which maps the street addresses to latitude and longitude, and then combine this information with high-resolution imagery. The resulting integration would enable applications where one can view an image and the system can identify the particular buildings in the image, including the business or occupant name, address, and telephone numbers. Similar integration can be done with other types of open source data. For example, we can extract train schedules for countries throughout the world and combine that with the vector data on railroads to predict what trains will pass on a particular line at a given time or to identify a specific train at a given point.

Applications Supported

As part of the DARPA Active Templates project, we have been working closely with Special Operations to automatically build Target Intelligence Packages from open source data and NIMA products. The equipment purchased is used to host all of our NIMA products. As part of this project, we received roughly two terabytes of geospatial data

products that are now available on our servers. Previously this data was only available in a storage room on individual CDs. The HeraclesMaps application that we built can now provide rapid access to geospatial data for any location in the world. The figure below shows a single screen shot from Heracles Maps showing an image, map, and vector data for the coastal area of Umm Qasr in Iraq.



Personnel Associated with the Research Effort:

(Note that since this is an equipment grant, no funds from the grant were used to support any of the personnel listed below. These individuals were supported under related projects.)

Craig Knoblock, PI
Cyrus Shahabi, Co-PI
Snehal Thakkar, GRA
Ching-Chien Chen, GRA
Jose Luis Ambite, Research Scientist
Maria Muslea, Research Programmer

Publications:

Ching-Chien Chen, Snehal Thakkar, Craig A. Knoblock, and Cyrus Shahabi.
Automatically Annotating and Integrating Spatial Datasets,
Volume 2750 of Lecture Notes in Computer Science, Springer, Berlin, 2003.

Interactions/Transitions:

We developed an application called HeraclesMaps, which is currently undergoing evaluation by SOCPAC (Special Operations in the Pacific). The application integrates a wide variety of geospatial data sources for use in planning and execution of operations. HeraclesMaps was recently tested in the CobraGold exercises that took place in Thailand in July, 2003 and received very positive reviews. The hardware purchased under this grant hosts image, map, vector, elevation, and point data that covers the world and is used to provide the data for the demonstration and evaluation of HeraclesMaps. Once this tool is deployed it will be used with classified data sources, but the availability of the hardware purchased under this grant has played a critical role in the transition of the technology from the research lab into actual use.

Discoveries/Inventions/Patent Disclosures

- None

Honors/Awards

- Invited talk by the PI at the Summer Assembly of the University Consortium for Geographic Information Science on Integrating Online and Geospatial Information Sources, Monterey, CA, June 17, 2003